Multi-Criteria Seed Selection for Targeted Influence Maximization within Social Networks

Artur Karczmarczyk[1[0000−0002−1135−7510]], Jarosław Jankowski[1[0000−0002−3658−3039]], and Jarosław Watrobski[2[0000−0002−4415−9414]]

[1] Faculty of Computer Science and Information Technology, West Pomeranian University of Technology in Szczecin, Żołnierska 49, 71-210 Szczecin
{artur.karczmarczyk,jaroslaw.jankowski}@zut.edu.pl

[2] The Faculty of Economics, Finance and Management of the University of Szczecin, Mickewicza 64, 71-101 Szczecin, Poland jwatrobski@usz.edu.pl

Abstract. Information spreading and influence maximization in social networks attracts attention from researchers from various disciplines. Majority of the existing studies focus on maximizing global coverage in the social network through initial seeds selection. In reality, networks are heterogeneous and different nodes can be a goal depending on campaign objectives. In this paper a novel approach with multi-attribute targeted influence maximization is proposed. The approach uses the multi-attribute nature of the network nodes (age, gender etc.) to better target specified groups of users. The proposed approach is verified on a real network and compared to the classic approaches delivers 7.14% coverage increase.

Keywords: seed selection · targeted influence maximization · MCDA.

1 Introduction

Social media are used for maintaining connections with relatives, friends and to access information sources. Virtual marketing within social media is strategized to reach people with specific interests. It results in a better engagement of the potential client thereof [4] and makes possible to avoid targeting users not interested in products or services. While most of the research focused on influence maximization and global coverage, social networking platforms deliver the ability to pick multiple choice parameters for an exact target class. The need to better address the real specifics of campaigns is visible, but the targeted approaches are introduced in a limited number of studies and are focused mainly on single node attributes [7] [15].

The approach presented in this paper deals with the selection of nodes for seeding the social platform on the basis of manifold criteria, as well as diverse attributes within agent based computational environment. The MCDA foundations of the proposed approach enable to adjust the gravity of each touchstone to be computed for selection purpose, in order to meet the requirements of the advertiser. Moreover, the relevant MCDA tools and computations enable to gauge

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

the impact of nodes seeding individually on the viral marketing strategy to hit the target groups. The paper comprises of five sections. The Introduction is followed by the Literature review section 2. Next, the methodology discussion is presented 3. After that, experimental results are showed 4 and followed by concluding statements 5.

2 Literature Review

In the area of information spreading within social platforms, it was supposed in the early stages of research that all the nodes of a network carry the same level of inclination towards a promulgated product or service or any other content [6]. However, in reality more result-oriented campaigns allow multiple node behaviors to be taken into consideration and better nodes allocation [7]. Recent studies used the cost assignment to the user of the network combined with the user interest benefits [8]. The goal of nodes selection can be also avoidance of intense campaign with unnecessarily repeated messages [1]. Pasumarthi et al. identified a targeted influence maximization problem, introducing an objective functionality and a penalizing criterion for adopting non targeted nodes [9].

Recently, initial studies are held discussing the application of MCDA techniques in the areas related to social networking. TOPSIS [3] method is used by Yang et al., in SIR (Susceptible Infected Recovered) model for identification of influential nodes in complex network [13]. Entropy weight method is used to measure and set up the weight values [14]. For maximizing the coverage and reducing the overlap, TOPSIS method is used by Zarei et al., while a social network is being influenced [16]. PROMETHEE [4] method was used by Karczmarczyk et al., to evaluate the responsiveness of viral marketing campaigns within social networking portals and also for providing decision support in order to plan these campaigns [5].

Review of studies in the area of information spreading and influence maximization has shown that among large number of studies only a small chunk is targeting the most common problem such as reaching out the specific user with multiple characteristics. Most of the existing approaches behave mono-trait by addressing nodes as a single attribute. However, social networks generally identify the target groups relying on multiple parameters, such as gender, localization or age. This identifies a research gap for seed selection based on a multi-characteristic computation in order to target specific multi-attribute network nodes, which this paper addresses.

3 Methodology

The proposed methodology complements the widely-used Independent Cascade (IC) model for modeling the spread within the complex networks [6], by taking

[3] Technique for Order Preference by Similarity to Ideal Solution
[4] Preferences Ranking Organization METHod for Enrichment of Evaluations

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

into account the problem of reaching targeted multi-attribute nodes in social networks by the information propagation processes. In the proposed approach, it is assumed that the network nodes are characterized not only by the centrality relations between them and other nodes, but also by a set of custom attributes $C_1, C_2, \ldots, C_n$. The nodes can also be characterized by the computed attributes derived from the network characteristics and measures, such as degree. Last, but not least, additional attributes can be derived as a composite of the two aforementioned types of attributes, by computing centrality measures based on limited subsets of the nodes' neighbors. For example, if attribute C_i represented the degree of a node, i.e. the total count of its neighbors, the C_{i_1} could represent the count of its male neighbors.

The aim of the proposed methodological framework is to maximize the influence within the targeted group of multi-attribute network nodes. While other approaches focus on generating the ranking of seeds based on a single centrality measure, in the authors' proposed methodological framework, the seeds are selected based on multiple attributes. This allows to select seeds which might be worse at maximizing global influence in the network, but which are better at maximizing influence in the targeted group of multi-attribute network nodes.

The approach presented in this paper is based on the MCDA methodology foundations [11]. The assumed modeling goal is to reach only the targeted set of multi-attribute nodes, instead of maximixing global influence in the network. Based on the guidelines provided by [3], it was decided that the PROMETHEE II method is most suitable for the proposed approach. It is an MCDA method that uses pairwise comparison and outranking flows to produce a ranking of the best decision variants. In the proposed approach, PROMETHEE II is used to produce a multi-criteria ranking of the nodes in the network with the aim to shortlist the ones which have the best chances to maximize influence in the targeted group of multi-attribute nodes. A detailed description of the PROMETHEE methods can be found in [2]. The MCDA foundations of the proposed approach help maximizing influence in the targeted group of multi-attribute nodes by selecting the seeds which have the highest, according to the marketer, potential to reach the targeted nodes in the social network. Moreover, the use of tools such as GAIA visual aid allows to understand the preferences backing the actual seed selection, and provide feedback which allows to further iteratively improve the obtained solution.

4 Empirical Study

In order to illustrate the proposed approach, the empirical study with the use of agent based simulations was performed on a relatively small real network [10] with 143 vertices and 623 edges giving the ability of detailed multi-criteria analysis. The proposed approach is intended for networks whose nodes are described with multiple attributes. However, the publicly available network datasets predominantly consist only of information on their nodes and edges, without information on the node attributes. To overcome this problem, the node attributes

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

Table 1. Criteria used in the empirical research.

Criterion	Values	Criterion	Values
C1 degree	integer [1-42]	C5 age	1: 0-29, 2: 30-59, 3: over 60
C2 gender	1: male, 2: female	C6 deg. younger	integer [0-18]
C3 deg. male	integer [0-20]	C7 deg. medium	integer [0-15]
C4 deg. female	integer [0-22]	C8 deg. older	integer [0-9]

Table 2. Top 7 network nodes used as seeds in the empirical research. A - degree; B - betweenness, C - closeness, D - eigen centrality, E-G - the proposed multi-attribute approach

A	105	17	95	48	132	43	91	E	17	95	48	132	50	105	20
B	107	17	48	91	32	95	141	F	19	95	48	50	132	91	105
C	105	17	95	37	74	48	91	G	132	20	136	19	50	122	3
D	105	31	136	132	20	19	69								

were artificially overlaid over the network, following the attributes' distribution from demographic data. Two demographic attributes were overlaid on the network – gender and age. For illustrative purposes, the target for the viral marketing campaign was chosen for the empirical research. In this experiment, the male users from the youngest age group were targeted, which translates to 28 of all the 143 users of the network. In the proposed approach, the seeds are selected from the network based on multiple criteria. In the empirical research, apart from the two aforementioned demographic attributes, also the degree measure was taken into account, as well as 5 criteria based on a mix of the degree and the demographic measures. This resulted in a total of 8 seed evaluation criteria, which are presented in Table 1.

Initially, the classic single-metric approaches were tested on the network, to provide a benchmark for the proposed approach. Four centrality metrics (degree, closeness, betweenness and eigen centrality) were used individually to first rank all vertices in the network, and then select the top nodes as seeds. It was decided for the seeding fraction to be set to 0.05 (seven seeding nodes) and propagation probability to 0.10. Moreover, in order to allow repeatability of the experiment for seeds selected by each approach, 10 pre-defined scenarios were created, in which each node was assigned a pre-drawn weight. The seeds obtained from rankings based on each centrality measure, i.e. degree, betweenness, closeness and eigen centrality, are presented in Table 2A - 2D respectively. The averaged simulation results are presented in Table 3A - 3D.

In the next step of the empirical study, the authors' proposed approach was used to choose the seeds based on a multi-criteria ranking produced by the PROMETHEE II method. All eight criteria were taken into account. Initially, the usual preference function was used for comparing each vertex under all criteria. Also, all criteria were given an equal preference weight (see Table 4E). As a result, seven seeds were selected (see Table 2E). It can be noticed that the

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

Table 3. Aggregated results from the empirical study simulations

	Iterations	Infected	Coverage	Infected targeted	Coverage
A	6.6	41.2	0.2881	7.7	0.2750
B	6.1	33.7	0.2357	5.5	0.1964
C	6.2	39.2	0.2741	6.2	0.2214
D	6.5	34.3	0.2399	9.0	0.3214
E	5.9	40.6	0.2839	9.2	0.3286
F	6.0	40.7	0.2846	9.5	0.3393
G	6.4	30.1	0.2105	9.7	0.3464

Table 4. Utilized PROMETHEE II parameters

	Criteria	C1	C2	C3	C4	C5	C6	C7	C8
E	Weight	1	1	1	1	1	1	1	1
	Preference function	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Usual
F	Weight	1	1	1	1	1	1	1	1
	Preference function	Linear	Usual	Linear	Linear	Usual	Linear	Linear	Linear
	q; p	3; 9	1; 2	1; 4	1; 2	1; 2	1; 4	1; 3	1; 2
G	Weight	8.2	25.4	12.6	3.8	28.4	14	3.8	3.8
	Preference function	Linear	Usual	Linear	Linear	Usual	Linear	Linear	Linear
	q; p	3; 9	1; 2	1; 4	1; 2	1; 2	1; 4	1; 3	1; 2

produced seed set is considerably different than the ones produced by the classic approaches (compare with Table 2A-2D).

After the simulations were executed with the newly selected seeds, it was observed that averagely 40.6 network nodes were infected (0.2839 coverage, see Table 3E). It is a worse result than for the degree-based approach. What is important to note, however, is that averagely 9.2 targeted nodes were infected, i.e. 0.3286 targeted coverage, which was the best result so far.

One of the benefits of using the PROMETHEE methods is the possibility to adjust the preference function used in pairwise comparisons of the nodes under individual criteria. While the usual preference function provides a simple boolean answer for the pairwise comparison of gender (C2) and age (C5) criteria, in case of the criteria based on degree, usage of a linear preference function with indifference and preference thresholds can yield better results. Therefore, in the subsequent step of the empirical research, a linear preference function was applied to all degree-based criteria (see Table 4F). The change in the preference function resulted in a different set of seeds selected for simulations (see Table 2F). The averaged results from the simulations are presented in Table 3F. It can be observed, that both global and targeted coverage values improved slightly.

Depending on the target group, the marketer can decide that some criteria can better help to reach the target group than the other criteria. Therefore, the marketer can adjust the preference weights of each criterion. Before the last set of simulations in this empirical research, an expert knowledge was elicited from the marketer with the use of the Analytical Hierarchy Process (AHP) [12], to adjust the preference weights of all criteria. The elicited weights used in the final set of

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

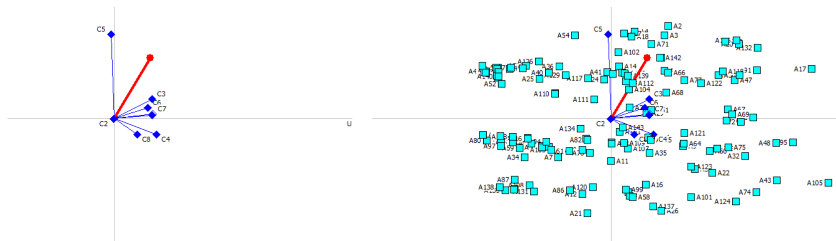

Fig. 1. GAIA Visual Analysis

simulations is presented in Table 4G. The adjusted preference weights resulted in a significantly different set of nodes used as seeds in the campaign (see Table 2G). The averaged simulation results are presented in Table 3G. The approach resulted in the best coverage in the targeted group (0.3464, compared to 0.2750 for the degree-based approach, 0.0714 difference). In the final step of the research, the GAIA visual analysis aid was used to study the criteria preference relations in the seed selection decision model (see Fig. 1). The analysis of Fig. 1 allows to observe that criteria C2 and C5 are not related to each other in terms of preference. This is quite straightforward, because these criteria represent the gender and age respectively. On the other hand, the remaining criteria are similar in terms of preference, possibly because they are all partially based on the degree measure.

5 Conclusions

The existing research in the area of information spreading focuses mainly on influence maximisation. Only limited number of studies discuss targeting nodes with specific characteristics with main focus on their single attributes. This paper proposes a novel approach to multi-attribute targeted influence maximization in social networks, focused on a multi-attribute seed selection. In the proposed approach, the seeds for initializing the campaign are chosen based on a ranking obtained with an MCDA method. The weights of individual criteria can be adjusted, as well as criteria values' comparison preference functions can be chosen to best fit the marketer's needs. In the experimental research, the proposed approach resulted in target nodes' coverage superior by as much as 7.14% compared to traditional degree-based approaches. The research opens some possible future directions. It would be beneficial to further broaden the research scope by studying how the changes in seeding fraction and propagation probability affect the efficiency of the proposed approach. Moreover, this research was performed on a network with attributes superimposed artificially. A research project can be run in order to collect knowledge about a real multi-attribute social network.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38

6 Acknowledgments

This work was supported by the National Science Centre of Poland, the decision no. 2017/27/B/HS4/01216 (AK, JJ) and within the framework of the program of the Minister of Science and Higher Education under the name "Regional Excellence Initiative" in the years 2019-2022, project number 001/RID/2018/19, the amount of financing PLN 10,684,000.00 (JW).

References

1. Abebe, R., Adamic, L., Kleinberg, J.: Mitigating overexposure in viral marketing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
2. Brans, J.P., Mareschal, B.: Promethee methods. In: Multiple criteria decision analysis: state of the art surveys, pp. 163–186. Springer (2005)
3. Cinelli, M., Kadziński, M., Gonzalez, M., Słowiński, R.: How to support the application of multiple criteria decision analysis? let us start with a comprehensive taxonomy. Omega p. 102261 (2020)
4. Iribarren, J.L., Moro, E.: Impact of human activity patterns on the dynamics of information diffusion. Phys. Rev. Lett. **103**, 038702 (Jul 2009)
5. Karczmarczyk, A., Jankowski, J., Watróbski, J.: Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks. PloS one **13**(12) (2018)
6. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146 (2003)
7. Mochalova, A., Nanopoulos, A.: A targeted approach to viral marketing. Electronic Commerce Research and Applications **13**(4), 283–294 (2014)
8. Nguyen, H.T., Dinh, T.N., Thai, M.T.: Cost-aware targeted viral marketing in billion-scale networks. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. pp. 1–9. IEEE (2016)
9. Pasumarthi, R., Narayanam, R., Ravindran, B.: Near optimal strategies for targeted marketing in social networks. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. pp. 1679–1680 (2015)
10. Rossi, R.A., Ahmed, N.K.: The network data repository with interactive graph analytics and visualization. In: AAAI (2015), http://networkrepository.com/email-enron-only.php
11. Roy, B., Vanderpooten, D.: The european school of MCDA: Emergence, basic features and current works **5**(1), 22–38
12. Saaty, T.L.: Decision-making with the ahp: Why is the principal eigenvector necessary. European journal of operational research **145**(1), 85–91 (2003)
13. Yang, P., Liu, X., Xu, G.: A dynamic weighted topsis method for identifying influential nodes in complex networks. Modern Physics Letters B **32**(19), 1850216 (2018)
14. Yang, Y., Yu, L., Zhou, Z., Chen, Y., Kou, T.: Node importance ranking in complex networks based on multicriteria decision making. Mathematical Problems in Engineering **2019** (2019)
15. Zareie, A., Sheikhahmadi, A., Jalili, M.: Identification of influential users in social networks based on users' interest. Information Sciences **493**, 217–231 (2019)
16. Zareie, A., Sheikhahmadi, A., Khamforoosh, K.: Influence maximization in social networks based on topsis. Expert Systems with Applications **108**, 96–107 (2018)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_38